Part III

Screening and Increasing Soluble Expression of Recombinant Proteins III

Application of Cell-free Expression Systems to Proteomic Studies 12

P.T. Beernink, B.W. Segelke, M.A.Coleman

Introduction

The promise of proteomics is to identify and characterize the physical and functional properties of proteins and protein complexes in parallel. Many proteomics efforts require the production of large numbers of purified proteins for biochemical or physical analyses. In particular, structural proteomics requires milligram quantities of highly purified proteins. However it has been established by pilot studies that soluble protein expression is one of the bottlenecks in structural proteomics processes (Fig. 12.1) [1]. Many proteins are inherently poorly expressed, insoluble, cytotoxic or subject to proteolysis, which results in low, soluble expression in vivo. Cell-free protein expression strategies can overcome some of these problems and yield a larger number of expressed proteins [2]. Cell-free expression can also be used to identify rapidly well-expressed proteins and to obtain proteins for biochemical and structural studies.

The speed of IVT expression is especially useful when modest quantities of protein are needed, for example in enzyme assays or microarray studies. Microarray-based methods represent a high-throughput approach to identify and characterize specific protein interactions. Microarrays are an alternative to the yeast two-hybrid screen [3], which is highly sensitive, but can give false results for misfolded or transactivating proteins. In addition, protein arrays can identify other interactions, such as those with DNA and small molecules, for unannotated proteins [4, 5]. Protein arrays allow multiplexed protein detection along with sensitive quantification in a dense format [6, 7]. Protein arrays also hold potential for miniaturization and portability and therefore have broad applications in basic research, genomic annotation, identification of disease markers and diagnosis of disease [7, 8, 9].

Methods

Template Preparation and IVT Reactions

Sequential PCR and IVT reactions were performed in 25-µl volumes in 96-well plates. PCR was performed using Taq DNA polymerase (Roche) and primers spe-

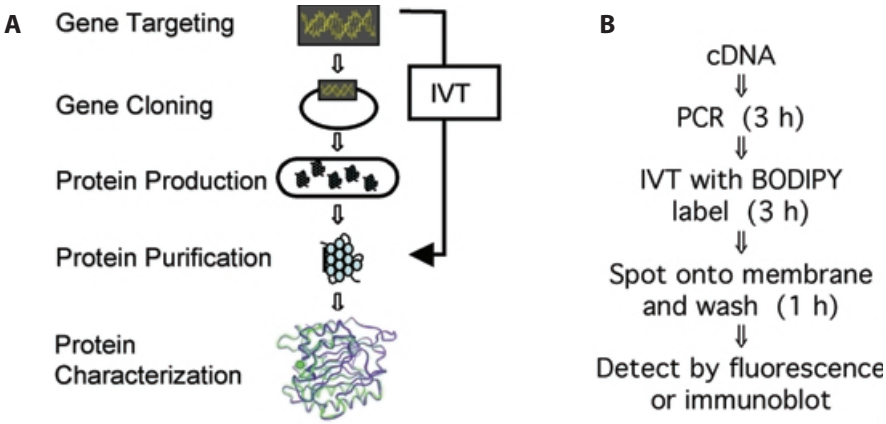

Fig. 12.1 A,B. Schemes for protein expression. **A** A typical scheme for protein production is shown, with protein expression being the largest bottleneck. *IVT* can circumvent cloning, bacterial expression and, for some applications, purification. **B** Scheme for linear template generation and cell-free protein production and detection. The process is automatable and can be performed in less than 8 h

cific to the T7 promoter GCGCGCGAGATCTCGATCCCGCGAAATTAATACGAC) and terminator (GCGCGCGTATCCGGATATAGTTCCTCCTTTCAG) sequences. RTS reactions (3 h, 30 °C) were carried out using 1 µl of unpurified PCR product and 0.13 µl of a BODIPY-Lys-tRNA conjugate (FluoroTect Green$_{Lys}$, Promega) for rapid visualization of the expressed protein. Expressed proteins were detected by a dot–blot procedure (5 µl IVT reaction) using an Immobilon-P membrane (Millipore) or by SDS-PAGE (6 µl, acetone-precipitated IVT reaction). Gels and blots were visualized on a FluorImager 595 (Molecular Devices) or by immunoblot using a Penta-His 1° antibody (Ab) (1:1000) (Qiagen) and an HRP-conjugated anti-mouse 2° Ab (Amersham).

Protein Microarrays

To analyze protein interactions, IVT-expressed proteins, purified proteins and antibodies (1–10 mg/ml) were spotted in duplicate on CMT-GAPS glass slides (Corning) with a robotic arrayer (Norgren Systems). Arrays contained up to 224 spots (~200 µm diameter) Controls included proteins (BSA, His$_6$-GFP, Ape1) and nucleic acids [M13 ssDNA, M13 dsDNA, 20-mer oligos, RNA, Cy-labelled DNA fragments (Molecular Probes)]. The arrays were dried at 25 °C and stored at 4 °C until use. Fluorescence was quantified using a ScanArray 5000 (Packard Bioscience) and visualized with false colour.

For far-Western experiments, arrays were incubated with 25–50 ng of a purified protein (15 min, 25 °C) and washed with phosphate-buffered saline (PBS) or PBS+ 0.1% Tween-20 (PBST). Interactions were detected with a 1° Ab

(RAD51 paralog- or SMARCAL1-specific Ab) (1:500), which was incubated (30 min, 25 °C) and washed with PBST. Rhodamine-labelled 2° Ab (1:250) was then added (15 min, 25 °C) and slides were washed and imaged as above.

Results

Choice of Expression System

We pursued cell-free protein expression in an *E. coli* extract because of its relatively high yield, its suitability for high-throughput automation and the potential to scale up reaction volumes (1–10 ml). We tested extracts from several suppliers and found that RTS extracts (Roche) provided favourable expression yields (data not shown). Using RTS extracts, we were able to perform efficient, cell-free expression from plasmid and PCR-amplified DNA templates.

To test the ability of IVT screening to predict bacterial expression levels, we examined the correlation of IVT- and *E. coli*-expressed proteins using 13 different human and bacterial clones, expressed as C-terminal green fluorescent protein (GFP) fusions [10]. The expression data cluster into two groups; the eight most highly expressed clones exhibit a good correlation (cc = 0.89), whereas the five least highly expressed clones display a weaker correlation (overall cc = 0.69). The in vitro expression levels of the latter five clones were significantly higher than the in vivo levels. Therefore, this set may comprise proteins that are cytotoxic or proteolytically sensitive, which underscores the benefits of cell-free expression for certain classes of proteins.

High-Throughput, Cell-Free Protein Expression

One useful application of high-throughput, cell-free protein expression is the identification of expressed proteins from hypothetical genes or cDNA expression libraries, which has previously been done in vivo [11]. We have applied the RTS system to high-throughput protein expression to identify highly expressed proteins for structural studies. This approach indicates which clones are candidates for larger RTS reactions that employ continuous nutrient exchange [12,13] to achieve yields of up to ~5 mg protein ml extract, or in bacterial expression systems.

Our screening strategy consists of several steps (Fig. 12.1), including: (1) PCR amplification of target genes; (2) cell-free protein expression using RTS 100, with optional incorporation of a fluorescent label; (3) transfer to membrane; and (4) detection by fluorescence or immunoblotting (see above Methods). Using fluorescence detection, the entire procedure can be carried out in approximately 7 h. For non-T7 based clones, an extra PCR amplification step can be performed to incorporate the necessary regulatory sequences (Linear Template Generation Set, Roche).

Applications of Cell-Free Expression

Protein Expression Screening

IVT expression testing was performed on 48 different clones, including prokaryotic and eukaryotic clones and several different expression plasmids, pIVEX2.4b (Roche), pET28 and pETBlue (Novagen). The blot was visualized by BODIPY fluorescence (Fig. 12.2A) and the spot intensities were quantified (Fig. 12.2B). This experiment illustrates that the BODIPY-Lys conjugate is efficiently incorporated, that there is wide variation among different constructs and that duplicate reactions are comparable. An immunoblot of the same

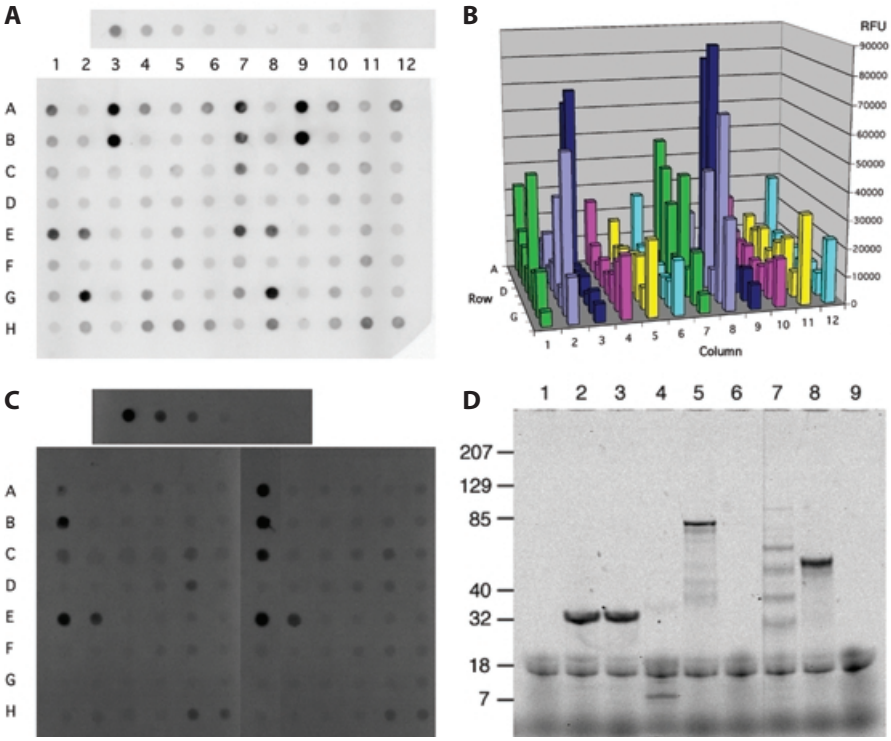

Fig. 12.2 A–D. High-throughput, cell-free protein expression. **A** BODIPY-labelled dot-blot of 48 clones, in duplicate (*columns 1–6 and 7–12*), imaged with a FluorImager 595 (Molecular Devices). Twofold serial dilutions of a His_6-GFP control reaction are shown in the *top strip*. **B** Quantification reported as relative fluorescence units (RFU) from blot in **A**. **C** Immunoblot of same membrane using a Penta-His Ab (Qiagen). Only a subset of clones encoded a His_6 tag. **D** SDS-PAGE of selected reactions from the same experiment. *Lane 1* Mass standards (at left kDa); *lane 2* purified GFP (3 μg); *lane 3* A7 (pIVEX-GFP); *lane 4* A8 (no DNA template); *lane 5* B7; *lane 6* C7; *lane 7* E7; *lane 8* G8; *lane 9* A9; *lane 10* B9

membrane using an affinity tag-specific antibody (Fig. 12.2C) offers the advantages that the signal does not depend on the number of lysine residues in the protein and that the quantities can be more easily standardized. In addition, proteins without BODIPY-conjugated lysine residues are more likely to retain their native functional properties. However, His_6 tags are also known to have effects on protein solubility and enzyme activity for some proteins [14]. The immunoblot (Fig. 12.2C) identifies some of the same clones as the BODIPY-based detection (A1, B1, C1, E1), but does not identify the clones lacking a His_6 tag (A3, B3, G2). Truncated proteins arising from premature translational termination are not detected in immunoblots of C-terminal His_6 tagged proteins.

Protein Microarrays

We have spotted IVT-expressed protein on microarrays for several applications. First, for expression screening, we have used arrays for comparison of relative expression levels. Control experiments were conducted to identify the limit of fluorescence detection (2×10^{-13} g) of purified GFP on a glass slide (CMT-GAPS, Corning) (Fig. 12.3A). IVT-expressed GFP fusion proteins were arrayed, which showed that spotting was reproducible and therefore useful to identify differences in relative expression levels (Fig. 12.3B). Second, for protein-specific detection, we have performed array-based immunoassays to detect GFP fusion proteins (Fig. 12.3C). These experiments can be adapted for protein expression profiling studies using cellular extracts. Third, for detection of protein interactions, an array-based far-Western technique has been developed. To investigate the interactions of putatively interacting proteins, a microarray was constructed that included nucleic acids, histones and nucleosomes. This array was used in far-Western experiments to identify interactions of SMARCAL1 [15] with nucleosomes and a Rad51 paralog with individual histones.

Biochemical Assays

It is important for subsequent biochemical studies that IVT-expressed proteins be functionally active. We have used IVT-expressed proteins to measure enzymatic activities and interactions. IVT- and bacterially expressed Ape1 enzymes were similar in DNA-binding and structure-specific nuclease activities (data not shown). IVT-expressed SMARCAL1 interacted with nucleosomes and not with individual histones H1, H2A, H2B and H4 (Fig. 12.3D). Also SMARCAL1–nucleosomal interactions were modulated in the presence of ATP. These results help to define the functional interactions of SMARCAL1. These and similar experiments have showed that proteins can be expressed in vitro and assayed directly for biochemical functions and interactions without purification.

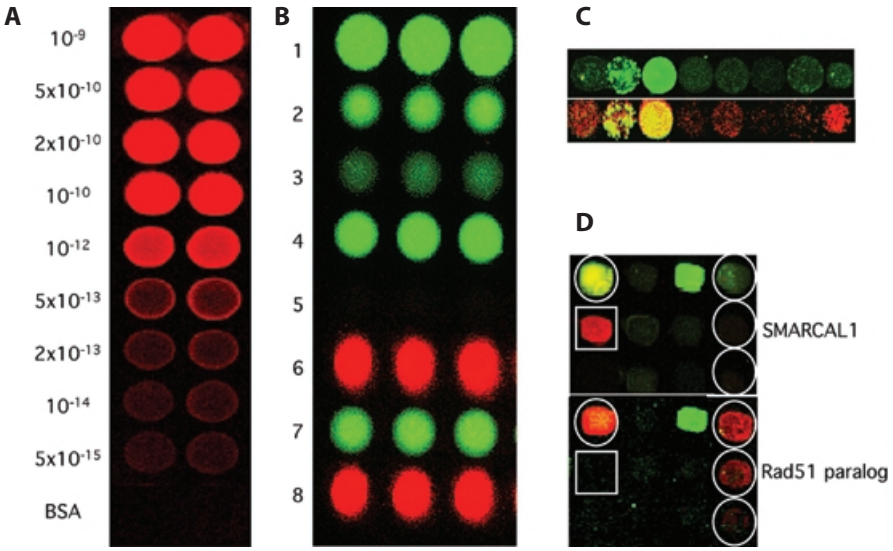

Fig. 12.3 A–D. Microarray-based protein visualization and interaction analyses. **A** Purified GFP spotted on a glass slide at various concentrations to determine detection sensitivity. **B** DNA (*red*) and IVT-expressed proteins (*green*) were arrayed on a glass slide to assess spotting reproducibility. *Row 1* LcrH; *2* GFP; *3* XRCC1; *4* LcrG; *5* IVT extract; *6* DNA; *7* SFN5; *8* DNA. **C** Array-based immunodetection of GFP fusion proteins. The *top row* shows the fluorescence of IVT-produced GFP fusions. The bottom row shows GFP fusion protein detection using an anti-GFP 1° Ab (BD Clontech) and a rhodamine-conjugated 2° antibody. **D** Protein interactions were demonstrated on an array containing 224 duplicate spots. SMARCAL1 associates primarily with nucleosomes (*squares*) and only weakly with individual histones (*circles*). In addition, a RAD51-paralog has affinity for free histones (*circles*) but not nucleosomes (*squares*)

Discussion

Our high-throughput expression and labelling system is rapid since it requires no subcloning or bacterial growth. Second, the detection system is flexible, since it can use antibodies for affinity-tagged proteins, or fluorescent labels or ^{35}S-Met for untagged proteins. Since BODIPY-Lys incorporation may affect protein conformation, immunodetection is preferable if the resulting protein is used for functional assays. BODIPY labelling is preferable if no tag or a variety of tags is present. Since the RTS is based on T7 expression, only the protein of interest and the tRNA conjugate are fluorescently labelled. Use of the RTS provides an alternative to in vivo screening methods [16]. This flexible approach for cell-free protein expression enables automated production of many proteins and their subsequent purification. The products are useful for downstream applications such as biochemical assays, protein arrays and in some cases structure determination.

Acknowledgements

We thank J. Albala, S. Doyle, M. Hadi, S. Krupka, V. Lao, B. Marsh, G. Martin, C. Prange, R. Rao, C. Sanders, J. Schafer and B. Souza for technical assistance and scientific discussions. This work was performed under the auspices of the US Dept. of Energy by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48 and was partially supported by a UC BioSTAR grant (01–10152) to P.T.B.

References

1. Pedelacq JD, Piltch E, Liong EC, Berendzen J, Kim CY, Rho BS, Park MS, Terwilliger TC, Waldo GS (2002) Engineering soluble proteins for structural genomics. Nat Biotechnol 20: 927–932
2. Yokoyama S (2003) Protein expression systems for structural genomics and proteomics. Curr Opin Chem Biol 7: 39–43
3. Fields S, Song O (1989) A novel genetic system to detect protein interaction. Nature 340: 245–246
4. Schweitzer B, Kingsmore SF (2002) Measuring proteins on microarrays. Curr Opin Biotechnol 13: 14–19
5. Pawlak M, Schick E, Bopp MA, Schneider MJ, Oroszlan P, Ehrat M (2002) Zeptosens' protein microarrays: a novel high-performance microarray platform for low abundance protein analysis. Proteomics 2: 383–393
6. Haab BB, Dunham MJ, Brown PO (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2:research0004.1-0004.13
7. Kononen J, Bubrndorf L, Kallioniemi A et al. (1998) Tissue microarrays for high throughput molecular profiling of tumor specimens. Nature Medicine 4: 844–847
8. Huang RP (2001) Detection of multiple proteins in an antibody-based protein microarray system. J Immunol Methods 255: 1–13
9. Wiese R, Belosludtsev Y, Powdrill T, Thompson P, Hogan M (2001) Simultaneous multianalyte ELISA performed on a microarray platform. Clinical Chem 47: 1451–1457
10. Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17: 691–695
11. Bussow K, Nordhoff E, Lubbert C, Lehrach H, Walter G (2000) A human cDNA library for high-throughput protein expression screening. Genomics 65: 1–8
12. Spirin AS (1992) Gene expression in cell-free systems on a preparative scale. Bioorg Khim 18: 1394–1402
13. Martin GA, Kawaguchi R., Lam Y, DeGiovanni A, Fukushima M, Mutter W (2001) High-yield, in vitro protein expression using a continuous-exchange, coupled transcription/translation system. BioTechniques 31: 948–953
14. Hammarstrom M, Hellgren N, van Den Berg S, Berglund H, Hard T (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci 11: 313–321
15. Coleman MA, Eisen JA, Mohrenweiser HW (2000) Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse. Genomics 65: 274–282
16. Doyle SA, Murphy MB, Massi JM, Richardson PM (2002) High-throughput proteomics: a flexible and efficient pipeline for protein production. J Proteome Res 1: 531–536